symmetry

MDPI

Article

A Three-Dimensional Subspace Algorithm Based on the Symmetry of the Approximation Model and WYL Conjugate Gradient Method

Guoxin Wang [1], Shengwei Yao [2], Mingyang Pei [1] and Jieqiong Xu [1,*]

[1] School of Mathematics and Information Science, Guangxi University, Nanning 530004, China; wgxin@st.gxu.edu.cn (G.W.); peimingyang@st.gxu.edu.cn (M.P.)
[2] Guangxi (ASEAN) Financial Research Center, Guangxi University of Finance and Economics, Nanning 530007, China; ysw@gxu.edu.cn
* Correspondence: xjq@gxu.edu.cn

Abstract: In this paper, a three-dimensional subspace method is proposed, in which the search direction is generated by minimizing the approximation model of the objective function in a three-dimensional subspace. The approximation model of the objective function is not unique, and alternatives can be chosen between a symmetric quadratic model and a conic model by specific criteria. Moreover, the idea of a WLY conjugate gradient method is applied to characterize the change of gradient direction between adjacent iteration points. The strategy of initial stepsize and nonmonotone line search are adopted, and the global convergence of the presented algorithm is established under mild assumptions. In numerical experiments, we use a collection of 80 unconstrained optimization test problems to show the competitive performance of the presented method.

Keywords: unconstrained optimization; subspace; conic model; global convergence

check for updates

Citation: Wang, G.; Yao, S.; Pei, M.; Xu, J. A Three-Dimensional Subspace Algorithm Based on the Symmetry of the Approximation Model and WYL Conjugate Gradient Method. *Symmetry* **2023**, *15*, 1207. https://doi.org/10.3390/sym15061207

Academic Editor: Guoqiang Wang

Received: 4 May 2023
Revised: 30 May 2023
Accepted: 1 June 2023
Published: 5 June 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Considering the following unconstrained optimization problem:

$$\min_{x\in\Re^n} f(x), \tag{1}$$

with an initial point x_0, the following iterative formula is often used to solve (1):

$$x_{k+1} = x_k + \alpha_k d_k, \tag{2}$$

where x_k is the kth iteration point, $\alpha_k \in \Re$ is the stepsize determined by a line search procedure, and d_k is the search direction obtained by special ways.

The search direction of the conjugate gradient (CG) methods is obtained by

$$d_{k+1} = -g_{k+1} + \beta_k d_k,\ k = 0, 1, ..., \tag{3}$$

where $d_0 = -g_0$, $g_{k+1} = \nabla f(x_{k+1})$ and $\beta_k \in \Re$ is a scalar called the conjugate gradient parameter. Corresponding to different choices for the parameter β_k, various nonlinear conjugate gradient methods have been proposed. Some classical CG methods include HS method (Hestenes and Stiefel [1]), FR method (Fletcher and Reeves [2]), PRP method (Polak et al. [3]), CD method (Fletcher [4]), LS method (Liu and Storey [5]) and DY method (Dai and Yuan [6]).

Among these mentioned CG methods, the PRP, HS and LS methods share the common numerator $g_{k+1}^T y_k$, where $y_k = g_{k+1} - g_k$. When the step $x_{k+1} - x_k$ is small, the factor y_k tends to zero, so β_k becomes small and the next search direction is essentially the steepest descent direction which can avoid jamming. Hence, these methods have better numerical performance than those lacking in $g_{k+1}^T y_k$. However, the PRP, HS and LS methods possess weak convergence properties. Based on the PRP method, Wei et al. [7] proposed a new formula

$$\beta_k^{WYL} = \frac{g_{k+1}^T y_k^*}{g_k^T g_k},$$

where $y_k^* = g_{k+1} - \frac{\|g_{k+1}\|}{\|g_k\|} g_k$ and $\|.\|$ denotes the Euclidean norm. When step $x_{k+1} - x_k$ is small, y_k^* also tends to zero so that the WYL method can also avoid jamming. Moreover, y_k^* can avoid the drawback: if the gradients of x_{k+1} and x_k are significantly different, the structure of y_k makes the smaller one between g_{k+1} and g_k nearly insignificant so that the method cannot make full use of the information. Furthermore, the WYL method not only has nice numerical performance but also has global convergence properties. More applications and properties of y_k^* can be found in [7–13].

The procedure for determining the stepsize α_k used in (2) is usually called line search, which can be classified into an exact one and an inexact one. One of the most used inexact line search is the so-called standard Wolfe line search [14]:

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k g_k^T d_k, \tag{4}$$

$$\nabla f(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k, \tag{5}$$

where $0 < \delta < \sigma < 1$. Obviously, it is a monotone procedure that seeks a suitable α_k, making the function value decrease to some extent. Zhang and Hager [15] proposed a nonmonotone version (ZH line search) that modifies condition (4) to

$$f(x_k + \alpha_k d_k) \leq C_k + \delta \alpha_k g_k^T d_k, \tag{6}$$

where $C_0 = f(x_0)$, $Q_0 = 1$, C_{k+1} and Q_{k+1} are updated by

$$C_{k+1} = \frac{\eta_k Q_k C_k + f(x_{k+1})}{Q_{k+1}}, \quad Q_{k+1} = \eta_k Q_k + 1, \tag{7}$$

where $\eta_k \in [\eta_{min}, \eta_{max}]$ and $0 \leq \eta_{min} \leq \eta_{max} \leq 1$. The choice of η_k controls the degree of nonmonotonicity. Such a line search can not only overcome some drawbacks in the monotone line search, but is particularly efficient for unconstrained problems in numerical experiments [15]. In addition to ZH line search, there are many other efficient nonmonotone line search procedures, which can be found in [16–25].

The subspace technique is one of the effective means for solving large-scale optimization problems, which is receivinging more and more attention. Yuan reviewed various subspace techniques that have been used in constructing numerical methods for solving nonlinear optimization problems in [26,27]. There are many optimization methods adopting the subspace technique, such as the limited-memory quasi-Newton method [28], the subspace trust region methods [29,30], the subspace SQP method [31], and more research can be found in [32–38]. Moreover, the combination between subspace technique and conjugate gradient method has been extensively studied. In the earliest research (see [39]), Yuan and Stoer computed the search direction d_{k+1} by minimizing the approximation quadratic model in the two dimensional subspace spanned by g_{k+1} and s_k, namely $\Omega_{k+1} = Span\{g_{k+1}, s_k\}$ where $s_k = x_{k+1} - x_k$, and proposed the subspace minimization conjugate gradient method (SMCG), in which d_{k+1} is formed by

$$d_{k+1} = t g_{k+1} + \mu s_k, \tag{8}$$

where t and μ are undetermined parameters. Based on the above idea, Andrei [40] extended the subspace to $\Omega_{k+1} = Span\{g_{k+1}, s_k, y_k\}$ and exploited the acceleration scheme, finally presenting a three-term conjugate gradient method (TTS), in which

$$d_{k+1} = -g_{k+1} + \mu s_k + \nu y_k,$$

and μ, ν are also scalar parameters. Inspired by Andrei, Yang et al. [41] changed the subspace into $\Omega_{k+1} = Span\{g_{k+1}, s_k, s_{k-1}\}$, and put forward the subspace three-term conjugate gradient method (STT). For the same subspace, Li et al. [42] added more parameters to the computation of search direction so that

$$d_{k+1} = tg_{k+1} + \mu s_k + \nu s_{k-1},$$

and adopted the strategy of nonmonotone line search, eventually proposing the subspace minimization conjugate gradient method with nonmonotone line search (SMCG_NLS). Huo et al. [13] combined the idea of the WYL method with subspace method, then constructed the subspace $\Omega_{k+1} = Span\{g_{k+1}, s_k, y_k^*\}$, in which

$$d_{k+1} = tg_{k+1} + \mu s_k + \nu y_k^*,$$

and finally proposed a new three-dimensional subspace conjugate gradient method (TSCG).

On the other hand, Dai and Kou [43] also focused on the analysis of Yuan and Stoer, but they paid more attention to the estimate of the parameter $\rho_{k+1} = g_{k+1}^T B_{k+1} g_{k+1}$ during the calculation of d_{k+1}. They combined the Barzilai–Borwein [44] idea and provided some efficient Barzilai–Borwein conjugate gradient methods (BBCG). It is remarkable that the idea of BBCG to estimate ρ_{k+1} is employed in this paper.

It is noteworthy that all of the above mentioned subspace minimization conjugate gradient methods obtain the search direction by minimizing the approximate quadratic model of objective function in the presented subspace. However, Sun and Yuan [45,46] have pointed out that, when the current iterative point is not close to the minimizer, the quadratic model may lead to a poor prediction of the minimizer if the objective function possesses strong non-quadratic behavior. Furthermore, a quadratic model does not take into account more information instead of the gradient value in the current iteration, which means that it does not have enough degrees freedom for incorporating all of the information in the iterative procedure.

Thus, the research for approximate nonquadratic model is of the essence. Up to now, many nonquadratic models have been applied to optimization problems, such as conic model, tensor model and regularization model. The conic model can incorporate more function information than the quadratic model, and its application in unconstrained optimization was first studied by Davidon [47]. A typical conic model for unconstrained optimization is

$$\phi_{k+1}(s) = \frac{g_{k+1}^T s}{1 + b_{k+1}^T s} + \frac{1}{2}\frac{s^T B_{k+1} s}{(1 + b_{k+1}^T s)^2},$$

which is an approximation to $f(x_k + s) - f(x_k)$, and B_{k+1} is a symmetric positive definite matrix approximating to the Hessian of $f(x)$ at x_{k+1} which satisfies the secant equation $B_{k+1}s_k = y_k$. The vector b_{k+1} is normally called the horizontal vector satisfying $1 + b_{k+1}^T s > 0$. Such a conic model has been investigated by many scholars. Sorensen [48] discussed a class of conic methods called "optimization by collinear scaling" for unconstrained optimization and shown that a particular member of this algorithm class has a Q-superlinear convergence. Ariyawansa [49] modified the procedure of Sorensen [48] and established the duality between the collinear scaling DFP and BFGS methods. Sheng [50] further discussed the interpolation properties of conic model method. Di and Sun [51] proposed a trust region method for conic models to solve unconstrained optimization problems. The trust region methods based on conic model have brought about a great number of research.

Li et al. [52] paid attention to the combination of subspace method and conic model. They considered the following conic approximation model:

$$\phi_{k+1}(d) = \frac{g_{k+1}^T d}{1 + b_{k+1}^T d} + \frac{1}{2}\frac{d^T B_{k+1} d}{(1 + b_{k+1}^T d)^2} \tag{9}$$

where

$$\begin{aligned} b_{k+1} &= -\frac{1 - \gamma_{k+1}}{\gamma_{k+1} g_{k+1}^T s_k} g_{k+1}, \\ \gamma_{k+1} &= \frac{-g_k^T s_k}{\sqrt{\Delta_{k+1}} + f_k - f_{k-1}}, \\ \Delta_{k+1} &= (f_k - f_{k+1})^2 - (g_{k+1}^T s_k)(g_k^T s_k). \end{aligned}$$

In the two-dimensional subspace $\Omega_{k+1} = Span\{g_{k+1}, s_k\}$, they minimized the above conic model to compute the search direction, and finally developed a subspace minimization conjugate method based on the conic model (SMCG_Conic). Sun et al. [53] extended the subspace to $\Omega_{k+1} = Span\{g_{k+1}, s_k, s_{k-1}\}$ and presented a three-dimensional subspace minimization conjugate gradient method based on a conic model (CONIC_CG3).

The adoption of a nonquadratic model does not mean abandoning the quadratic model; here we refer to the analysis of Yuan [54], in which a quantity u_k is defined by

$$u_k = \left| \frac{2(f_k - f_{k+1} + g_{k+1}^T s_k)}{s_k^T y_k} - 1 \right|,$$

which shows the extent of how the objective function $f(x)$ is close to a quadratic on the line segment between x_k and x_{k+1}. Dai et al. [55] indicate that if the following condition

$$u_k \leq c_1 \; or \; \max\{u_k, u_{k-1}\} \leq c_2 \tag{10}$$

holds, where $0 < c_1 < c_2$ are two small constants, then they think that $f(x)$ is very close to a quadratic on the line segment between x_k and x_{k+1}. If the condition (10) is satisfied, then the choice of the quadratic approximation model is preferable; otherwise, the conic model is more suitable. The utilization of such a quantity can be found in [56,57].

Since the combination of subspace technique and conic model has been investigated, we come up with the idea of whether we can make further studies. SMCG_Conic and CONIC_CG3 are efficient methods that possess good theoretical and numerical results, but they only considered the two-dimensional subspace $Span\{g_{k+1}, s_k\}$ and three-dimensional subspace $Span\{g_{k+1}, s_k, s_{k-1}\}$, respectively. Hence, we expect to extend the application of conic model in the subspace conjugate gradient method. In addition, the WYL method with the ingredient y_k^* has good performance both in theory and numerical experiment, and the subspace method TSCG based on quadratic model and subspace $Span\{g_{k+1}, s_k, y_k^*\}$ also is an efficient method. In this paper, we investigates a new application of WYL method and subspace method based on the conic model by extending SMCG_Conic to subspace $Span\{g_{k+1}, s_k, y_k^*\}$. Furthermore, we adopt the initial stepsize strategy and nonmonotone line search, under which we establish the sufficient descent property and global convergence property. Since our method is based on three-dimensional subspace and conic model, it has efficient performance in numerical experiment and can solve more problems than subspace method based on two-dimensional subspace and quadratic model.

This paper is organized as follows: in Section 2, the search directions on the subspace Ω_{k+1} based on two different models are derived, and the criteria for how to choose the approximate model and search direction are presented. In Section 3, we obtain the stepsize by the strategies of initial stepsize and nonmonotone line search, and elaborate the generated algorithm. In Section 4, we give the proofs for some important lemmas of the search direction, then state the convergence performance of the generated algorithm under suitable assumptions. In Section 5, we compare the numerical results of our algorithm with another two methods.

2. The Search Direction

For an optimization method, the search direction attaches great importance to the iterative formula, and the computation of search direction is always the first job. Hence, the main content of this section is to construct the formula for the search direction. Since the approximation model is selected between conic model and quadratic model, the computation will be divided into two parts.

2.1. Conic Model

In this subsection, we will give the computation of the search direction in the case that conic model is used.

When (10) is violated, the conic model is more suitable to approximate the objective function, so we compute the search direction by minimizing the conic model in subspace $\Omega_{k+1} = Span\{g_{k+1}, s_k, y_k^*\}$. We consider the subproblem

$$\min_{d \in \Omega_{k+1}} \phi_{k+1}(d), \tag{11}$$

where $\phi_{k+1}(d)$ is the same as (9).

The discussion will be divided into three parts depending on the three different dimensions of the subspace Ω_{k+1}.

Situation 1: $dim(\Omega_{k+1}) = 3$.

In this situation, the search direction is

$$d_{k+1} = tg_{k+1} + \mu s_k + \nu y_k^*. \tag{12}$$

By substituting (12) into (11) and using the secant equation, the problem (11) turns into

$$\min_{(t,\mu,\nu)} \phi_{k+1}(u) = \frac{a^T u}{1 + c^T u} + \frac{1}{2} \frac{u^T A_{k+1} u}{(1 + c^T u)^2}, \tag{13}$$

where

$$A_{k+1} = \begin{pmatrix} \rho_{k+1} & g_{k+1}^T y_k & \omega_k \\ g_{k+1}^T y_k & s_k^T y_k & y_k^T y_k^* \\ \omega_k & y_k^T y_k^* & \tau_k \end{pmatrix}, \ a = \begin{pmatrix} \|g_{k+1}\|^2 \\ g_{k+1}^T s_k \\ g_{k+1}^T y_k^* \end{pmatrix}, \ c = \begin{pmatrix} b_{k+1}^T g_{k+1} \\ b_{k+1}^T s_k \\ b_{k+1}^T y_k^* \end{pmatrix}, \ u = \begin{pmatrix} t \\ \mu \\ \nu \end{pmatrix},$$

and $\rho_{k+1} = g_{k+1}^T B_{k+1} g_{k+1}$, $\omega_k = g_{k+1}^T B_{k+1} y_k^*$, $\tau_k = {y_k^*}^T B_{k+1} y_k^*$.

To solve (13), we first figure out the zero solution of the function

$$\nabla \phi_{k+1}(u) = \frac{1}{1 + c^T u} \left(I - \frac{cu^T}{1 + c^T u} \right) \left(a + \frac{A_{k+1} u}{1 + c^T u} \right),$$

and obviously $I - \frac{cu^T}{1+c^T u}$ is invertible, so the problem is reduced to $a + \frac{A_{k+1} u}{1+c^T u} = 0$. After calculation, we can deduce that, when A_{k+1} is positive definite and $1 + c^T A_{k+1}^{-1} a \neq 0$, the solution of (13) is

$$u_{k+1} = \frac{-A_{k+1}^{-1} a}{1 + c^T A_{k+1}^{-1} a}. \tag{14}$$

By using the relationship $A_{k+1}^{-1} = \frac{A_{k+1}^*}{|A_{k+1}|}$, where

$$A_{k+1}^* = \begin{pmatrix} X & \theta_1 & \theta_2 \\ \theta_1 & \theta & \theta_3 \\ \theta_2 & \theta_3 & Y \end{pmatrix} \tag{15}$$

is the adjugate matrix of A_{k+1}, and X, θ_1, θ_2, θ, θ_3, Y denote the cofactor of A_{k+1}^{11}, A_{k+1}^{21}, A_{k+1}^{31}, A_{k+1}^{22}, A_{k+1}^{32}, A_{k+1}^{33}, respectively, in which A_{k+1}^{ij} denotes the ij-th element of A_{k+1} and

$$\begin{aligned} X &= (s_k^T y_k)\tau_k - (y_k^T y_k^*)^2, \\ \theta_1 &= (y_k^T y_k^*)\omega_k - (g_{k+1}^T y_k)\tau_k, \\ \theta_2 &= (g_{k+1}^T y_k)(y_k^T y_k^*) - (s_k^T y_k)\omega_k, \\ \theta &= \rho_{k+1}\tau_k - \omega_k^2, \\ \theta_3 &= (g_{k+1}^T y_k)\omega_k - \rho_{k+1}(y_k^T y_k^*), \\ Y &= \rho_{k+1}(s_k^T y_k) - (g_{k+1}^T y_k)^2, \end{aligned}$$

then we finally obtain the solution of (13),

$$u_{k+1} = \begin{pmatrix} t_{k+1} \\ \mu_{k+1} \\ \nu_{k+1} \end{pmatrix} = -\frac{1}{D_{k+1}} \begin{pmatrix} X & \theta_1 & \theta_2 \\ \theta_1 & \theta & \theta_3 \\ \theta_2 & \theta_3 & Y \end{pmatrix} \begin{pmatrix} \|g_{k+1}\|^2 \\ g_{k+1}^T s_k \\ g_{k+1}^T y_k^* \end{pmatrix} = \frac{1}{D_{k+1}} \begin{pmatrix} -q_1 \\ -q_2 \\ -q_3 \end{pmatrix}, \tag{16}$$

where

$$\begin{aligned} |A_{k+1}| &= \rho_{k+1} X + \theta_1 g_{k+1}^T y_k + \theta_2 \omega_k, \\ D_{k+1} &= |A_{k+1}| + q_1 b_{k+1}^T g_{k+1} + q_2 b_{k+1}^T s_k + q_3 b_{k+1}^T y_k^*, \\ q_1 &= X\|g_{k+1}\|^2 + \theta_1 g_{k+1}^T s_k + \theta_2 g_{k+1}^T y_k^*, \\ q_2 &= \theta_1 \|g_{k+1}\|^2 + \theta g_{k+1}^T s_k + \theta_3 g_{k+1}^T y_k^*, \\ q_3 &= \theta_2 \|g_{k+1}\|^2 + \theta_3 g_{k+1}^T s_k + Y g_{k+1}^T y_k^*. \end{aligned}$$

So the solution of (11) in the three-dimensional subspace $\Omega_{k+1} = Span\{g_{k+1}, s_k, y_k^*\}$ is

$$d_{k+1} = \begin{pmatrix} g_{k+1} & s_k & y_k^* \end{pmatrix} u_{k+1}.$$

During evaluation, it is important to find appropriate ways to estimate ρ_{k+1}, ω_k and τ_k in order to avoid matrix–vector multiplication and improve efficiency. Yuan and Stoer [39] proposed two ways to calculate such quantities containing B_{k+1}, one of which is to obtain B_{k+1} by using the scaled memoryless BFGS formula. The approach of Dai and Kou [43] is to combine the Barzilai–Borwein [44] idea by approximating the Hessian by $(1/\alpha_{k+1}^{BB_1})I$ or $(1/\alpha_{k+1}^{BB_2})I$, where

$$\alpha_{k+1}^{BB_1} = \frac{\|s_k\|^2}{s_k^T y_k}, \quad \alpha_{k+1}^{BB_2} = \frac{s_k^T y_k}{\|y_k\|^2}.$$

For ω_k, we utilize the memoryless BFGS formula to obtain B_{k+1} so that

$$\begin{aligned} \omega_k &= g_{k+1}^T \left(I + \frac{y_k y_k^T}{s_k^T y_k} - \frac{s_k s_k^T}{s_k^T s_k} \right) y_k^* \\ &= g_{k+1}^T y_k^* + \frac{g_{k+1}^T y_k y_k^T y_k^*}{s_k^T y_k} - \frac{g_{k+1}^T s_k s_k^T y_k^*}{s_k^T s_k}. \end{aligned} \tag{17}$$

Then for τ_k, we combine the idea of [42,43], and estimate τ_k by

$$\tau_k = \zeta_k \frac{\|y_k\|^2}{s_k^T y_k} \|y_k^*\|^2, \tag{18}$$

where

$$\zeta_k = \begin{cases} \max\{0.9\zeta_{k-1}, 1.2\}, & \text{if } \alpha_k > 1.0, \\ \min\{1.1\zeta_{k-1}, 1.75\}, & \text{otherwise}, \end{cases}$$

and $\zeta_0 = 1.5$. It is obvious that $\zeta_k \in [1.2, 1.75]$.

For the estimate of ρ_{k+1}, we adopt the idea of Li et al. [42], because it can guarantee some good properties. Firstly, according to the analysis regarding (14), the positive definitiveness of A_{k+1} is essential, which requires $|A_{k+1}| > 0$. It follows that

$$\rho_{k+1} > \frac{-\theta_1 g_{k+1}^T y_k - \theta_2 \omega_k}{X}. \tag{19}$$

By setting $X = m_k s_k^T y_k \tau_k$ with $m_k \triangleq 1 - \frac{(y_k^T y_k^*)^2}{s_k^T y_k \tau_k}$, and combining (15), we have

$$\rho_{k+1} > \left[\frac{(g_{k+1}^T y_k)^2}{s_k^T y_k} + \frac{\omega_k^2}{\tau_k} - 2\frac{g_{k+1}^T y_k \omega_k y_k^T y_k^*}{s_k^T y_k \tau_k} \right] / m_k \triangleq n_k, \tag{20}$$

if and only if m_k is positive, which can be guaranteed by (18). In addition, the positive definitiveness of A_{k+1} also requires that its first and second order leading principal minors are positive, and it means

$$\rho_{k+1} > \frac{(g_{k+1}^T y_k)^2}{s_k^T y_k}. \tag{21}$$

Secondly, $D_{k+1} > 0$ is also a necessary condition to keep the sufficient descent property, from which follows

$$\rho_{k+1} > \frac{S_k}{s_k^T y_k \tau_k} / M_k, \tag{22}$$

if $M_k > 0$, where

$$\begin{aligned} S_k = -\theta_1 g_{k+1}^T y_k - \theta_2 \omega_k &- b_{k+1}^T g_{k+1} (X\|g_{k+1}\|^2 + \theta_1 g_{k+1}^T s_k + \theta_2 g_{k+1}^T y_k^*) \\ &- b_{k+1}^T s_k (\theta_1 \|g_{k+1}\|^2 - \omega_k^2 g_{k+1}^T s_k + \omega_k g_{k+1}^T y_k g_{k+1}^T y_k^*) \\ &- b_{k+1}^T y_k^* (\theta_2 \|g_{k+1}\|^2 + \omega_k g_{k+1}^T y_k g_{k+1}^T s_k - (g_{k+1}^T y_k)^2 g_{k+1}^T y_k^*), \end{aligned}$$

and

$$M_k = 1 - \frac{(y_k^T y_k^*)^2}{s_k^T y_k \tau_k} + \frac{1 - \gamma_{k+1}}{\gamma_{k+1}} \left[2\frac{g_{k+1}^T y_k^* y_k^T y_k^*}{s_k^T y_k \tau_k} - \frac{g_{k+1}^T s_k}{s_k^T y_k} - \frac{(g_{k+1}^T y_k^*)^2}{g_{k+1}^T s_k \tau_k} \right].$$

Likewise, we define the right-hand side of (22) as N_k, i.e. $N_k \triangleq \frac{S_k}{s_k^T y_k \tau_k} / M_k$.

After the above discussion, we can estimate ρ_{k+1} as follows,

$$\rho_{k+1} = \zeta_k \max\{K, N_k, n_k\}, \tag{23}$$

where $K = K_1 \|g_{k+1}\|^2$,

$$K_1 = \max\left\{ \frac{\|y_k\|^2}{s_k^T y_k}, \left| \frac{1 - \gamma_{k+1}}{\gamma_{k+1}} \right| \frac{\|g_{k+1}\|^2}{|g_{k+1}^T s_k|} \right\},$$

and ζ_k is the same as that in (18). One of the purposes of K is to ensure (21).

Before computing d_{k+1} by (12) and (16), we should verify the following conditions:

$$\Delta_{k+1} \geq 0, \tag{24}$$

$$M_k \geq \rho_0, \tag{25}$$

$$\xi_1 \leq \frac{s_k^T y_k}{\|s_k\|^2} \leq \frac{\|y_k\|^2}{s_k^T y_k} \leq \xi_2, \tag{26}$$

$$\left|\frac{1-\gamma_{k+1}}{\gamma_{k+1}}\right| \frac{\|g_{k+1}\|^2}{|g_{k+1}^T s_k|} \leq \xi_3, \tag{27}$$

$$\frac{|A_{k+1}|}{s_k^T y_k \tau_k \rho_{k+1}} \geq \xi_5, \tag{28}$$

where $\rho_0 \in (0,1)$ and $\xi_1, \xi_2, \xi_3, \xi_5$ are positive constants. The expressions (24) and (25) are fundamental premises of the conic model (5) and relation (22), respectively. On the basis of the Barzilai–Borwein [44] idea, (26) might indicate the suitable condition numbers of the approximation Hessian matrix. Moreover, (27) is vital to guarantee the descent property of the search direction. As for (28), obviously it makes A_{k+1} more positive definite, and is also important for establishing the sufficient descent property of the search direction.

For convenience, we call (24)–(28) the first conic model conditions. Hence, if the first conic model conditions hold, then we compute the search direction by (12) and (16).

Situation 2: $dim(\Omega_{k+1}) = 2$ or 1.

Li et al. [52] have made a deep study of the subspace conjugate gradient method based on the conic model in this case. Here we refer to their works. When $dim(\Omega_{k+1}) = 2$, the search direction is formed by

$$d_{k+1} = t g_{k+1} + \mu s_k,$$

which is the same as (8). The solution of t and μ is

$$\begin{pmatrix} t_{k+1} \\ \mu_{k+1} \end{pmatrix} = \frac{1}{\bar{D}_{k+1}} \begin{pmatrix} -c_1 \\ -c_2 \end{pmatrix}, \tag{29}$$

which is the same as (13) in [52].

In this case, conditions (24) and (26) are still essential, but before obtaining d_{k+1} by (8) and (29), the following conditions are also necessary:

$$\bar{m}_k \geq \bar{\rho}_0, \tag{30}$$

$$\left|\frac{1-\gamma_{k+1}}{\gamma_{k+1}}\right| \frac{\|g_{k+1}\|^2}{|g_{k+1}^T s_k|} \leq \xi_3, \text{ if } \frac{1-\gamma_{k+1}}{\gamma_{k+1} g_{k+1}^T s_k} < 0, \tag{31}$$

$$\frac{\|g_{k+1}\|^2 \|y_k\| \|s_k\|}{(g_{k+1}^T s_k)^2} \leq \xi_4, \text{ if } \frac{1-\gamma_{k+1}}{\gamma_{k+1} g_{k+1}^T s_k} \geq 0, \tag{32}$$

where $\bar{\rho}_0 \in (0,1)$, ξ_4 is a positive constant and ξ_3 is identical to that in (27). We call (24), (26) and (30)–(32) the second conic model conditions. If the first conic model conditions do not all hold but the second conic model conditions hold, we compute the search direction by (8) and (29).

If the first and second conic model conditions are violated but (33) and (34) hold,

$$\vartheta_1 \leq \frac{s_k^T y_k}{\|s_k\|^2}, \tag{33}$$

$$\frac{|g_{k+1}^T y_k g_{k+1}^T d_k|}{d_k^T y_k \|g_{k+1}\|^2} \leq \vartheta_4, \tag{34}$$

where $\vartheta_4 \in [0,1)$ and ϑ_1 are positive constants, the HS direction is considered, i.e.,

$$d_{k+1} = -g_{k+1} + \beta_k^{HS} d_k. \tag{35}$$

We call (33) and (34) the HS conditions.

Therefore, there are two choices of search direction when $dim(\Omega_{k+1}) = 2$: one is to compute d_{k+1} by (8) and (29) if the second conic model conditions hold; the other is to compute d_{k+1} by (35) if the HS conditions hold. Otherwise, we use the negative gradient direction $-g_{k+1}$ as our search direction.

2.2. Quadratic Model

The content of this subsection is to calculate the search direction in the case that the objective function is approximated by a quadratic model.

When the condition (10) holds, the quadratic approximation model seems to perform well in approximating the objective function, so we consider the subproblem based on the quadratic approximation model as follows:

$$\min_{d \in \Omega_{k+1}} \psi_{k+1}(d) = g_{k+1}^T d + \frac{1}{2} d^T B_{k+1} d, \tag{36}$$

and related work can be found in [13]. The corresponding results can be applied to this situation. We state the concrete results in the following.

Firstly, there are some necessary conditions:

$$\vartheta_1 \le \frac{s_k^T y_k}{\|s_k\|^2} \le \frac{\|y_k\|^2}{s_k^T y_k} \le \vartheta_2, \tag{37}$$

$$\vartheta_1 \le \frac{\tau_k}{\|y_k^*\|^2}, \frac{4\|y_k\|^4 \|y_k^*\|^2}{(s_k^T y_k)^2 \tau_k} \le \vartheta_2, \tag{38}$$

$$\frac{\|s_k\|^2}{\|g_{k+1}\|^2} \ge \vartheta_3, \tag{39}$$

we call (37)–(39) the quadratic model conditions.

When the quadratic model conditions hold,

$$d_{k+1} = t_{k+1}^1 g_{k+1} + \mu_{k+1}^1 s_k + \nu_{k+1}^1 y_k^*, \tag{40}$$

where

$$\begin{pmatrix} t_{k+1}^1 \\ \mu_{k+1}^1 \\ \nu_{k+1}^1 \end{pmatrix} = -\frac{1}{|A_{k+1}|} \begin{pmatrix} X & \theta_1 & \theta_2 \\ \theta_1 & \theta & \theta_3 \\ \theta_2 & \theta_3 & \Upsilon \end{pmatrix} \begin{pmatrix} \|g_{k+1}\|^2 \\ g_{k+1}^T s_k \\ g_{k+1}^T y_k^* \end{pmatrix}.$$

When the quadratic model conditions do not all hold but (37) holds,

$$d_{k+1} = t_{k+1}^2 g_{k+1} + \mu_{k+1}^2 s_k, \tag{41}$$

where

$$\begin{pmatrix} t_{k+1}^2 \\ \mu_{k+1}^2 \end{pmatrix} = \frac{1}{|\bar{A}_{k+1}|} \begin{pmatrix} g_{k+1}^T y_k g_{k+1}^T s_k - \|g_{k+1}\|^2 s_k^T y_k \\ \|g_{k+1}\|^2 g_{k+1}^T y_k - \rho_{k+1} g_{k+1}^T s_k \end{pmatrix}.$$

$$\bar{A}_{k+1} = \begin{pmatrix} \rho_{k+1} & g_{k+1}^T y_k \\ g_{k+1}^T y_k & s_k^T y_k \end{pmatrix}.$$

When (37) does not holds but the HS conditions hold,

$$d_{k+1} = -g_{k+1} + \beta_k^{HS} d_k.$$

Otherwise, $d_{k+1} = -g_{k+1}$.

To sum up, the generated directions possess several forms as follows:
when the approximation conic model is selected, i.e. (10) is violated,
- d_{k+1} is calculated by (12) and (16), if the first conic model conditions hold,
- d_{k+1} is calculated by (8) and (29), if the second conic model conditions hold,
- d_{k+1} is calculated by (35), if the HS conditions hold;

when the approximation quadratic model is selected, i.e. (10) holds,
- d_{k+1} is calculated by (12) and (40), if the quadratic model conditions hold,
- d_{k+1} is calculated by (8) and (41), if only (37) holds,
- d_{k+1} is calculated by (35), if the HS conditions hold;

otherwise, $d_{k+1} = -g_{k+1}$.

3. The Stepsize and Algorithm

In this section, we will present the obtainment of another essential ingredient in the iterative formula of optimization method, i.e. stepsize. It requires two steps: the first is the choice of initial stepsize, and the second is the line search procedure. After these discussion, then we can propose our whole algorithm.

3.1. Strategy for the Initial Stepsize

Firstly, we show how to choose the initial stepsize.

It is generally acknowledged that the initial stepsize is of great significance for an optimization method, especially for the conjugate gradient method. For Newton and quasi-Newton methods, the choice of the initial trial stepsize may always be unit step $\alpha^0 = 1$. However, it is different for methods that do not produce well-scaled search directions, such as the steepest descent or the conjugate gradient methods. Thus, it is significant to make a reasonable initial guess of the stepsize by considering the current information about the objective function and algorithm for such methods [58,59]. Many strategies of the initial stepsize have been proposed which can be found in the achievements of Hager and Zhang [60], Nocedal and Wright [58], Dai and Kou [19], Liu and Liu [57].

In the strategy of [19], Dai and Kou presented a condition,

$$\frac{|\varphi_{k+1}(\alpha_{k+1}^0) - \varphi_{k+1}(0)|}{\varepsilon_1 + |\varphi_{k+1}(0)|} \leq \varepsilon_2, \tag{42}$$

where $\varepsilon_1, \varepsilon_2$ are small positive constants, α_{k+1}^0 denotes the initial trial stepsize, and $\varphi_{k+1}(\alpha) = f(x_{k+1} + \alpha d_{k+1})$. If condition (42) holds, it implies that the points $x_{k+1} + \alpha_{k+1}^0 d_{k+1}$ and x_{k+1} are not far away from each other, so it is reasonable to use the minimizer of the quadratic interpolation function for $\varphi_{k+1}(0)$, $\varphi_{k+1}^{'}(0)$ and $\varphi_{k+1}(\alpha_{k+1}^0)$ as the new initial stepsize, where $\varphi_{k+1}^{'}(0)$ denotes the first derivative of $\varphi_{k+1}(0)$.

In this paper, the selection of the initial stepsize has two parts, depending on whether the negative gradient direction is adopted or not, and is presented via modification of that in Liu and Liu [56].

When the search direction is negative gradient direction, according to the analysis of Liu and Liu [57], it is desirable to take the initial trial stepsize by

$$\bar{\bar{\alpha}}_{k+1} = \begin{cases} \max\{\min\{s_k^T y_k / \|y_k\|^2, \lambda_{max}\}, \lambda_{min}\}, & \text{if } g_{k+1}^T s_k > 0, \\ \max\{\min\{\|s_k\|^2 / s_k^T y_k, \lambda_{max}\}, \lambda_{min}\}, & \text{if } g_{k+1}^T s_k \leq 0, \end{cases} \tag{43}$$

where λ_{min} and λ_{max} are two positive constants controlling the initial stepsize within the interval $[\lambda_{min}, \lambda_{max}]$ which is preferable in numerical experiments.

Andrei [61] thinks that, the more accurate the step length is, the faster convergence a conjugate gradient algorithm has. It therefore makes sense to verify if the initial trial stepsize $\bar{\bar{\alpha}}_{k+1}$ satisfies (42) or not. If so and $d_k \neq -g_k$, $\|g_{k+1}\|^2 \leq 1$, we update the initial stepsize by

$$\tilde{\alpha}_{k+1} = \max\{\min\{\tilde{\tilde{\alpha}}_{k+1}, \lambda_{max}\}, \lambda_{min}\},$$

where

$$\tilde{\tilde{\alpha}}_{k+1} = \min q(\varphi_{k+1}(0), \varphi_{k+1}^{'}(0), \varphi_{k+1}(\bar{\bar{\alpha}}_{k+1})),$$

where $q(\varphi_{k+1}(0), \varphi_{k+1}^{'}(0), \varphi_{k+1}(\bar{\bar{\alpha}}_{k+1}))$ denotes the quadratic interpolation function for $\varphi_{k+1}(0)$, $\varphi_{k+1}^{'}(0)$ and $\varphi_{k+1}(\bar{\bar{\alpha}}_{k+1})$.

Therefore, the initial stepsize for the negative gradient direction is

$$\alpha_{k+1}^{0} = \begin{cases} \tilde{\alpha}_{k+1}, & \text{if (42) holds, } d_k \neq -g_k,\ \|g_{k+1}\|^2 < 1 \text{ and } \tilde{\tilde{\alpha}}_{k+1} > 0, \\ \bar{\bar{\alpha}}_{k+1}, & \text{otherwise.} \end{cases} \tag{44}$$

When it comes to the search direction that is not negative gradient direction, the similarity of the calculation between our direction and that in quasi-Newton methods implies that the unit stepsize $\alpha_{k+1}^{0} = 1$ might be a reasonable initial trial stepsize. Again we figure out the minimizer of the quadratic interpolation function

$$\bar{\alpha}_{k+1} = \min q(\varphi_{k+1}(0), \varphi_{k+1}^{'}(0), \varphi_{k+1}(1)).$$

If $\alpha_{k+1}^{0} = 1$ satisfies (42) and $\bar{\alpha}_{k+1} > 0$, we update the initial stepsize by

$$\hat{\alpha}_{k+1} = \max\{\min\{\bar{\alpha}_{k+1}, \lambda_{max}\}, \lambda_{min}\}.$$

Therefore, the initial stepsize for the search direction except negative gradient direction is

$$\alpha_{k+1}^{0} = \begin{cases} \hat{\alpha}_{k+1}, & \text{if (42) holds and } \bar{\alpha}_{k+1} > 0, \\ 1, & \text{otherwise.} \end{cases} \tag{45}$$

3.2. The Nonmonotone Line Search

Secondly, we introduce the line search used in our algorithm.

For the variable η_k in ZH line search, Zhang and Hager [15] proved that if $\eta_{max} = 1$, then the generated sequence $\{x_k\}$ only has the property that

$$\liminf_{k\to\infty} \|g_k\| = 0.$$

Liu and Liu [57] presented a formula of η_k,

$$\eta_k = \begin{cases} c, & \bmod\ (k,n) = n-1, \\ 1, & \bmod\ (k,n) \neq n-1, \end{cases}$$

where $0 < c < 1$ and $\mathrm{mod}(k,n)$ denotes the residue for k modulo n, and it resulted in a better convergence.

Referring to the above study, Li et al. [52] and Sun et al. [53] made some modifications so that the improved line search is more appropriate for their algorithms.

In order to gain the decent convergence result and performance, this paper modifies the improved ZH line search used in [53]. To be specific, we set

$$C_{k+1} = \begin{cases} f_{k+1} + \min\{1,\ 0.9(C_k - f_{k+1})\}, & k < 5, \\ [\eta_k Q_k C_k + f(x_{k+1})]/Q_{k+1}, & k \geq 5, \end{cases} \tag{46}$$

$$Q_{k+1} = \begin{cases} 6, & k = 4, \\ \eta_k Q_k + 1, & k \geq 5, \end{cases} \tag{47}$$

where

$$\eta_k = \begin{cases} \eta, & \bmod\ (k, n) = 0, \\ 1, & \bmod\ (k, n) \neq 0, \end{cases} \tag{48}$$

where $\eta = 0.999$.

3.3. Algorithm

In this subsection, we will detail our new three-dimensional subspace method based on conic model.

First, we incorporate a special restart technique proposed by Dai and Kou [19]. They defined a quantity

$$r_k = \frac{2(f_{k+1} - f_k)}{\alpha_k (g_k^T d_k + g_{k+1}^T d_k)},$$

where $f_{k+1} = \varphi_k(\alpha_k)$ and $f_k = \varphi_k(0)$. If r_k is close to 1, they think that the line search function φ_k is close to some quadratic function. According to their analysis, the exact approach of this quantity is that if there are continuously many iterations such that r_k is close to 1, we restart the algorithm with the negative gradient direction. In addition, if the number of the iterations since the last restart reaches the MaxRestart threshold, we also restart out algorithm.

The details of the three-dimensional subspace conjugate gradient method based on conic model are given in the next algorithm, which is Algorithm 1.

Algorithm 1 WYL_TSCO

```
Require: initial point $x_0$, initial stepsize $\alpha_0^0$, positive constants $\epsilon$, $\epsilon_1$, $\epsilon_2$, $0 < \delta < \sigma < 1$,
    $\xi_1, \xi_2, \xi_3, \xi_4, \xi_5, \vartheta_1, \vartheta_2, \vartheta_3, \vartheta_4, \rho_0, \bar{\rho}_0 \in (0,1), \lambda_1, \lambda_2, \lambda_{min}, \lambda_{max}$
Ensure: optimal $x^*$
1: Set MaxRestart: = 4n, IterRestart: = 0, IterQuad: = 0, MinQuad: = 3, Numnongrad: = 0, $C_0 = f_0$, $d_0 = -g_0$
   and $k$: = 0.
2: If $\|g_0\|_\infty \leq \epsilon$, stop.
3: Calculate the stepsize $\alpha_k$ by (5) and (6) with $\alpha_k^0$.
4: Update $x_{k+1} = x_k + \alpha_k d_k$. If $\|g_{k+1}\|_\infty \leq \epsilon$, stop; otherwise, let IterRestart: = IterRestart +1. If $|r_k - 1| \leq 10^{-8}$
   or $|f_{k+1} - f_k - 0.5(g_{k+1}^T s_k + g_k^T s_k)| \leq 6.0 \times 10^{-8}$, IterQuad: = IterQuad+1; else, IterQuad: = 0.
5: Calculate the search direction $d_k$. If Numnongrad = MaxRestart, or (IterQuad = MinQuad and IterRestart
   $\neq$IterQuad), go to 5.6; else if the condition (10) holds, go to 5.1; else, go to 5.3.
    5.1: If the quadratic model conditions hold, compute $d_{k+1}$ by (12) and (40), set Numnongrad: = Numnongrad+1; else, go to 5.2.
    5.2: If condition (37) holds, compute $d_{k+1}$ by (8) and (41), set Numnongrad: = Numnongrad+1; else, go
         to 5.5.
    5.3: If the first conic model conditions hold, compute $d_{k+1}$ by (12) and (16), set Numnongrad: = Numnongrad+1; else, go to 5.4.
    5.4: If the second conic model conditions, compute $d_{k+1}$ by (8) and (29), set Numnongrad: = Numnongrad+1; else, go to 5.5.
    5.5: If the HS conditions hold, compute $d_{k+1}$ by (35), set Numnongrad: = Numnongrad +1; else, go to 5.6.
    5.6: Compute $d_{k+1} = -g_{k+1}$, set Numnongrad: = 0 and IterRestart: = 0.
6: If $d_{k+1} = -g_{k+1}$, calculate $\alpha_{k+1}^0$ by (44); otherwise, calculate $\alpha_{k+1}^0$ by (45).
7: Update $Q_{k+1}$ and $C_{k+1}$ by (46)–(48).
8: Set $k := k + 1$, go to Step 3.
```

4. Properties of the Proposed Algorithm

The theoretical nature is an important criterion to determine the performance of an algorithm. In this section, we will give some properties of WYL_TSCO, including the sufficient descent property and the global convergence.

4.1. Some Lemmas

This subsection will make a deep discussion on our algorithm and prove some properties of the presented directions before establishing the global convergence. At first, we propose two assumptions as follows:

Assumption 1. *The objective function $f(x)$ is continuously differentiable and bounded from below on $\Re^n$.*

Assumption 2. *The gradient function $g(x)$ is Lipschitz continuous on the level set $D = \{x \in \Re^n : f(x) \le f(x_0)\}$, which means that there exists a positive constant $L > 0$ satisfying*

$$\|g(x) - g(y)\| \le L\|x - y\|,\ \forall x, y \in D,$$

which implies that $\|y_k\| \le L\|s_k\|$.

With these assumptions, we can prove some properties of the new directions in the following.

Lemma 1. *For the search direction d_{k+1} calculated by WYL_TSCO, there exists a constant $\varsigma_1 > 0$ such that*

$$g_{k+1}^T d_{k+1} \le -\varsigma_1 \|g_{k+1}\|^2. \tag{49}$$

Proof. We will discuss in several parts based on different situations and different approximation models.

Case I: if the negative direction is adopted, i.e. $d_{k+1} = -g_{k+1}$, then

$$g_{k+1}^T d_{k+1} = -\|g_{k+1}\|^2 \le -\frac{1}{2}\|g_{k+1}\|^2.$$

Case II: if d_{k+1} is determined by (35), i.e., the HS direction, combining (34), then we have

$$\begin{aligned} g_{k+1}^T d_{k+1} &= -\|g_{k+1}\|^2 + \beta_k^{HS} g_{k+1}^T d_k \\ &\le -\|g_{k+1}\|^2 + \frac{|g_{k+1}^T y_k g_{k+1}^T d_k|}{d_k^T y_k} \\ &\le -\|g_{k+1}\|^2 + \vartheta_4 \|g_{k+1}\|^2 \\ &= -(1 - \vartheta_4)\|g_{k+1}\|^2. \end{aligned}$$

Case III (conic): if d_{k+1} is calculated by (8) and (29), Li, Liu and Liu [52] have proved that

$$g_{k+1}^T d_{k+1} \le -\bar{\varsigma}\|g_{k+1}\|^2,$$

where

$$\bar{\varsigma} = \min\left\{ \frac{\bar{\rho}_0}{(8 - 6\bar{\rho}_0 \max\{\xi_2, \xi_4\})}, \frac{1}{6\bar{c}} \right\},$$

and $\bar{c}$ is a positive constant.

Case IV (conic): when d_{k+1} is obtained by (12) and (16), we have

$$\begin{aligned} g_{k+1}^T d_{k+1} &= \begin{pmatrix} \|g_{k+1}\|^2 \\ g_{k+1}^T s_k \\ g_{k+1}^T y_k^* \end{pmatrix}^T \begin{pmatrix} t_{k+1} \\ \mu_{k+1} \\ \nu_{k+1} \end{pmatrix} \\ &= -\frac{1}{D_{k+1}} \begin{pmatrix} \|g_{k+1}\|^2 \\ g_{k+1}^T s_k \\ g_{k+1}^T y_k^* \end{pmatrix}^T \begin{pmatrix} X & \theta_1 & \theta_2 \\ \theta_1 & \theta & \theta_3 \\ \theta_2 & \theta_3 & Y \end{pmatrix} \begin{pmatrix} \|g_{k+1}\|^2 \\ g_{k+1}^T s_k \\ g_{k+1}^T y_k^* \end{pmatrix} \\ &= -\frac{\|g_{k+1}\|^4}{D_{k+1}} h(x,y), \end{aligned}$$

where $x \triangleq \frac{g_{k+1}^T y_k^*}{\|g_{k+1}\|^2}$, $y \triangleq \frac{g_{k+1}^T s_k}{\|g_{k+1}\|^2}$. Moreover, $h(x,y)$ is a binary quadratic function of x and y which can be expressed as

$$h(x,y) = Yx^2 + 2\theta_3 xy + \theta y^2 + 2\theta_2 x + 2\theta_1 y + X.$$

It is easy to acquire the Hessian of $h(x,Y)$

$$H_h = \begin{pmatrix} 2Y & 2\theta_3 \\ 2\theta_3 & 2\theta \end{pmatrix},$$

we have $Y > 0$ because A_{k+1} is positive definite, and the determinant of H_h

$$4Y\theta - 4\theta_3^2 = 4\rho_{k+1}|A_{k+1}|,$$

is also positive, so $h(x,y)$ has a minimizer, that is

$$h(x,y)_{min} = \frac{|A_{k+1}|}{\rho_{k+1}}.$$

Therefore, we can derive

$$g_{k+1}^T d_{k+1} \leq -\frac{\|g_{k+1}\|^4}{D_{k+1}} h(x,y)_{min} \leq -\frac{|A_{k+1}|}{D_{k+1}\rho_{k+1}} \|g_{k+1}\|^4. \tag{50}$$

Because ρ_{k+1}, D_{k+1} and $|A_{k+1}|$ are all positive, we just need to prove that $\frac{|A_{k+1}|}{D_{k+1}\rho_{k+1}} \|g_{k+1}\|^2$ has a lower bound. Since D_{k+1} contains ρ_{k+1}, we first prove that ρ_{k+1} has an upper bound; that is the upper bound of N_k, n_k and K.

For N_k, we use the Cauchy inequality and can obtain

$$\begin{aligned} |N_k| &= \left| \frac{S_k}{s_k^T y_k \tau_k} \right| / M_k \\ &\leq \Bigg[\left| \frac{1-\gamma_{k+1}}{\gamma_{k+1}} \right| \Bigg(\frac{\|g_{k+1}\|^4}{|g_{k+1}^T s_k|} + 4\frac{\|g_{k+1}\|^4 \|y_k\|^2 \|y_k^*\|^2}{|g_{k+1}^T s_k| s_k^T y_k \tau_k} + 4\frac{\|g_{k+1}\|^2 \|y_k\| \|y_k^*\| |\omega_k|}{s_k^T y_k \tau_k} \\ &\qquad + 2\frac{\|g_{k+1}\|^3 \|y_k\|}{s_k^T y_k} + 2\frac{\|g_{k+1}\|^3 \|y_k^*\| |\omega_k|}{|g_{k+1}^T s_k| \tau_k} + \frac{\|g_{k+1}\| \|s_k\| \omega_k^2}{s_k^T y_k \tau_k} \Bigg) \\ &\qquad + \frac{\|g_{k+1}\|^2 \|y_k\|^2}{s_k^T y_k} + \frac{\omega_k^2}{\tau_k} + 2\frac{\|g_{k+1}\| \|y_k\|^2 \|y_k^*\| |\omega_k|}{s_k^T y_k \tau_k} \Bigg] / M_k, \end{aligned}$$

combining (17), (18) and using the Cauchy inequality again, we have

$$\begin{aligned}|N_k| \le \frac{1}{M_k}\|g_{k+1}\|^2\Bigg[\left|\frac{1-\gamma_{k+1}}{\gamma_{k+1}}\right|\Bigg(&\frac{\|g_{k+1}\|^2}{|g_{k+1}^T s_k|}+4\frac{\|g_{k+1}\|^2}{|g_{k+1}^T s_k|}+4\frac{2+\xi_2}{\zeta_k}\frac{\|g_{k+1}\|}{\|y_k\|}\\ &+2\frac{\|g_{k+1}\|\|y_k\|}{s_k^T y_k}+2\frac{2+\xi_2}{\zeta_k}\frac{s_k^T y_k}{\|y_k\|^2}\frac{\|g_{k+1}\|^2}{|g_{k+1}^T s_k|}+\frac{(2+\xi_2)^2}{\zeta_k}\frac{\|g_{k+1}\|\|s_k\|}{\|y_k\|^2}\Bigg)\\ &+\frac{\|y_k\|^2}{s_k^T y_k}+\frac{(2+\xi_2)^2}{\zeta_k}\frac{s_k^T y_k}{\|y_k\|^2}+2\frac{2+\xi_2}{\zeta_k}\Bigg].\end{aligned}$$

Under condition (26), we have $\frac{\|y_k\|\|s_k\|}{s_k^T y_k} \le \sqrt{\frac{\xi_2}{\xi_1}}$, $\frac{\|s_k\|}{\|y_k\|} \le \frac{1}{\xi_1}$, and

$$\frac{\|g_{k+1}\|\|y_k\|}{s_k^T y_k} = \frac{\|g_{k+1}\|\|y_k\||g_{k+1}^T s_k|}{s_k^T y_k|g_{k+1}^T s_k|} \le \frac{\|y_k\|\|s_k\|}{s_k^T y_k}\frac{\|g_{k+1}\|^2}{|g_{k+1}^T s_k|},$$

and in the same way,

$$\frac{\|g_{k+1}\|}{\|y_k\|} \le \frac{\|s_k\|}{\|y_k\|}\frac{\|g_{k+1}\|^2}{|g_{k+1}^T s_k|}.$$

Note that $\zeta_k \ge 1$, the above inequality of $|N_k|$ can be simplified to

$$\begin{aligned}|N_k| \le \frac{1}{M_k}\|g_{k+1}\|^2\Bigg[\left|\frac{1-\gamma_{k+1}}{\gamma_{k+1}}\right|\Bigg(&5+4(2+\xi_2)\frac{\|s_k\|}{\|y_k\|}+2\frac{\|y_k\|\|s_k\|}{s_k^T y_k}+2(2+\xi_2)\frac{s_k^T y_k}{\|y_k\|^2}\\ &+(2+\xi_2)^2\frac{\|s_k\|^2}{\|y_k\|^2}\Bigg)\frac{\|g_{k+1}\|^2}{|g_{k+1}^T s_k|}+\left(1+(2+\xi_2)\frac{s_k^T y_k}{\|y_k\|^2}\right)^2\frac{\|y_k\|^2}{s_k^T y_k}\Bigg].\end{aligned}$$

Utilizing conditions (25)–(27) and the expression of K, we can obtain the upper bound of N_k,

$$\begin{aligned}|N_k| &\le \frac{\|g_{k+1}\|^2}{M_k}\Bigg[\left(\frac{(2+\xi_2)^2}{\xi_1^2}+6\frac{(2+\xi_2)}{\xi_1}+2\sqrt{\frac{\xi_2}{\xi_1}}+5\right)\left|\frac{1-\gamma_{k+1}}{\gamma_{k+1}}\right|\frac{\|g_{k+1}\|^2}{|g_{k+1}^T s_k|}\\ &\qquad+\left(\frac{2+\xi_2}{\xi_1}+1\right)^2\frac{\|y_k\|^2}{s_k^T y_k}\Bigg]\\ &\le \frac{1}{\rho_0}\left(2\frac{(2+\xi_2)^2}{\xi_1^2}+8\frac{(2+\xi_2)}{\xi_1}+2\sqrt{\frac{\xi_2}{\xi_1}}+6\right)K_1\|g_{k+1}\|^2\\ &= \left(2\frac{(2+\xi_2)^2}{\xi_1^2}+8\frac{(2+\xi_2)}{\xi_1}+2\sqrt{\frac{\xi_2}{\xi_1}}+6\right)\frac{K}{\rho_0}.\end{aligned}$$

Next, the upper bound of n_k is acquired by the same procedure,

$$
\begin{aligned}
|n_k| &\le \frac{1}{m_k}\|g_{k+1}\|^2\left(\frac{\|y_k\|^2}{s_k^T y_k} + \frac{(2+\xi_2)^2}{\zeta_k}\frac{s_k^T y_k}{\|y_k\|^2} + 2\frac{(2+\xi_2)}{\zeta_k}\right) \\
&\le \frac{1}{m_k}\|g_{k+1}\|^2\left(1+(2+\xi_2)\frac{s_k^T y_k}{\|y_k\|^2}\right)^2\frac{\|y_k\|^2}{s_k^T y_k} \\
&\le \frac{1}{1-\frac{1}{\zeta_k}}\frac{(\xi_1+\xi_2+2)^2}{\xi_1^2}K_1\|g_{k+1}\|^2 \\
&\le 6\frac{(\xi_1+\xi_2+2)^2}{\xi_1^2}K.
\end{aligned}
$$

For K_1, it is easy to know that

$$
K_1 \le \max\{\xi_2, \xi_3\}.
$$

Through the above discussion, now we can give the upper bound of ρ_{k+1},

$$
\begin{aligned}
\rho_{k+1} &= \zeta_k \max\{K, N_k, n_k\} \\
&\le 2\max\left\{\left(2\frac{(2+\xi_2)^2}{\xi_1^2} + 8\frac{(2+\xi_2)}{\xi_1} + 2\sqrt{\frac{\xi_2}{\xi_1}} + 6\right)/\rho_0,\ 6\frac{(\xi_1+\xi_2+2)^2}{\xi_1^2}\right\}K_1\|g_{k+1}\|^2 \\
&\le L_0\|g_{k+1}\|^2,
\end{aligned}
$$

where

$$
L_0 \triangleq 2\max\left\{\left(2\frac{(2+\xi_2)^2}{\xi_1^2} + 8\frac{(2+\xi_2)}{\xi_1} + 2\sqrt{\frac{\xi_2}{\xi_1}} + 6\right)/\rho_0,\ 6\frac{(\xi_1+\xi_2+2)^2}{\xi_1^2}\right\}\max\{\xi_2, \xi_3\}.
$$

Since we have found the upper bound of ρ_{k+1}, then we now consider D_{k+1}. According to (22), D_k can be expressed as

$$
D_{k+1} = s_k^T y_k \tau_k \rho_{k+1} M_k - S_k = s_k^T y_k \tau_k (M_k \rho_{k+1} - M_k N_k). \tag{51}
$$

Using the formula of M_k and the upper bound of ρ_{k+1}, we have

$$
\begin{aligned}
|D_{k+1}| = D_{k+1} \le \|g_{k+1}\|^2 s_k^T y_k \tau_k \Bigg[L_0 + L_0\frac{\|y_k\|^2\|y_k^*\|^2}{s_k^T y_k \tau_k} + L_0\left|\frac{1-\gamma_{k+1}}{\gamma_{k+1}}\right|\Bigg(2\frac{\|g_{k+1}\|\|y_k\|\|y_k^*\|^2}{s_k^T y_k \tau_k} \\
+ \frac{\|g_{k+1}\|\|s_k\|}{s_k^T y_k} + \frac{\|g_{k+1}\|^2\|y_k^*\|^2}{|g_{k+1}s_k|\tau_k}\Bigg) + \frac{M_k|N_k|}{\|g_{k+1}\|^2}\Bigg].
\end{aligned}
$$

Note that

$$
\frac{\|g_{k+1}\|\|y_k\|\|y_k^*\|^2}{s_k^T y_k \tau_k} = \frac{\|g_{k+1}\|\|y_k\|\|y_k^*\|^2|g_{k+1}s_k|}{s_k^T y_k \tau_k |g_{k+1}s_k|} \le \frac{\|s_k\|}{\|y_k\|}\frac{\|g_{k+1}\|^2}{|g_{k+1}s_k|},
$$

$$
\frac{\|g_{k+1}\|\|s_k\|}{s_k^T y_k} = \frac{\|g_{k+1}\|\|s_k\||g_{k+1}s_k|}{s_k^T y_k |g_{k+1}s_k|} \le \frac{\|s_k\|^2}{s_k^T y_k}\frac{\|g_{k+1}\|^2}{|g_{k+1}s_k|},
$$

so the above inequality of D_{k+1} can be simplified to

$$D_{k+1} \le \|g_{k+1}\|^2 s_k^T y_k \tau_k \left[2L_0 + L_0 \left(2\frac{\|s_k\|}{\|y_k\|} + \frac{\|s_k\|^2}{s_k^T y_k} + \frac{s_k^T y_k}{\|y_k\|^2} \right) \left| \frac{1-\gamma_{k+1}}{\gamma_{k+1}} \right| \frac{\|g_{k+1}\|^2}{|g_{k+1} s_k|} \right. \\ \left. + \left(4\frac{\xi_2}{\xi_1} + 6\sqrt{\frac{\xi_2}{\xi_1}} + 16 \right) K_1 \right].$$

Finally, using (26) and (27), we can obtain

$$\begin{aligned} D_{k+1} &\le \|g_{k+1}\|^2 s_k^T y_k \tau_k \left[2L_0 + \left(4\frac{L_0}{\xi_1} + 4\frac{\xi_2}{\xi_1} + 6\sqrt{\frac{\xi_2}{\xi_1}} + 16 \right) K_1 \right] \\ &\le \|g_{k+1}\|^2 s_k^T y_k \tau_k \left[2L_0 + \left(4\frac{L_0}{\xi_1} + 4\frac{\xi_2}{\xi_1} + 6\sqrt{\frac{\xi_2}{\xi_1}} + 16 \right) \max\{\xi_2, \xi_3\} \right] \\ &\le L_1 \|g_{k+1}\|^2 s_k^T y_k \tau_k, \end{aligned} \tag{52}$$

where

$$L_1 \triangleq \left[2L_0 + \left(4\frac{L_0}{\xi_1} + 4\frac{\xi_2}{\xi_1} + 6\sqrt{\frac{\xi_2}{\xi_1}} + 16 \right) \max\{\xi_2, \xi_3\} \right].$$

With (28), (50) and (52), the sufficient descent property of d_{k+1} under this case can be established by

$$g_{k+1}^T d_{k+1} \le -\frac{\xi_5}{L_1} \|g_{k+1}\|^2.$$

Case V (quadratic): when d_{k+1} is generated by (8) and (41), the proof can be found in [43], in which Dai and Kou proved that

$$g_{k+1}^T d_{k+1} \le -\frac{1}{2\vartheta_2} \|g_{k+1}\|^2. \tag{53}$$

Next, if d_{k+1} is generated by (12) and (40), Yang et al. [13] have drawn that

$$g_{k+1}^T d_{k+1} \le -\frac{1}{16\vartheta_2} \|g_{k+1}\|^2. \tag{54}$$

Combining (54) with (53), we finally have

$$g_{k+1}^T d_{k+1} \le -\frac{1}{16\vartheta_2} \|g_{k+1}\|^2.$$

After the above discussion, we can prove that there indeed exists a constant ς_1 such that

$$g_{k+1}^T d_{k+1} \le -\varsigma_1 \|g_{k+1}\|^2,$$

where

$$\varsigma_1 = \min\left\{ \bar{\varsigma}, \frac{1}{2}, 1-\vartheta_4, \frac{\xi_5}{L_1}, \frac{1}{16\vartheta_2} \right\}.$$

The proof is complete. □

Lemma 2. *Assume that f satisfies Assumption 2. If the search direction d_{k+1} is calculated by WYL_TSCO, then there exists a constant $\varsigma_2 > 0$ such that*

$$\|d_{k+1}\| \le \varsigma_2 \|g_{k+1}\|. \tag{55}$$

Proof. Similar to Lemma 1, the proof is divided into several cases.

Case I: if d_{k+1} is calculated by negative gradient, then there certainly holds

$$\|d_{k+1}\| \leq \|g_{k+1}\|.$$

Case II: if the search direction is the HS direction, according to Assumption 2 and the condition (33), we have

$$\begin{aligned} \|d_{k+1}\| &= \| - g_{k+1} + \beta_k^{HS} d_k\| \\ &\leq \|g_{k+1}\| + \frac{\|g_{k+1}\|\|y_k\|\|d_k\|}{d_k^T y_k} \\ &\leq \left(1 + \frac{L}{\vartheta_1}\right)\|g_{k+1}\|. \end{aligned}$$

Case III (conic): if d_{k+1} is calculated by (8) and (29), from [52] we conclude that

$$\|d_{k+1}\| \leq \left(\frac{10\xi_2 + 5\xi_1 + 5m}{\bar{\rho}_0 \xi_1^2}\right)\|g_{k+1}\|,$$

where $m = 2n_0\bar{\xi}/\bar{\rho}_0$ with

$$n_0 = \max\left\{4 - 3\bar{\rho}_0, 1 + \frac{2\xi_2}{\xi_1}\right\},\ \bar{\xi} = \max\{\xi_2, \xi_3, \xi_4\}.$$

Case IV (conic): if d_{k+1} is formed by (12) and (16), we have

$$\begin{aligned} \|d_{k+1}\| &= \|t_{k+1}g_{k+1} + \mu_{k+1}s_k + \nu_{k+1}y_k^*\| \\ &\leq \frac{1}{D_{k+1}}(|q_1|\|g_{k+1}\| + |q_2|\|s_k\| + |q_3|\|y_k^*\|), \end{aligned} \tag{56}$$

so in order to prove Lemma 2 in this case, we first need to obtain a lower bound of D_{k+1}. Combining (23), (25) and (51), and the value range of ζ_k, we can derive

$$\begin{aligned} D_{k+1} &= s_k^T y_k \tau_k M_k(\zeta_k \max\{K, N_k, n_k\} - N_k) \\ &\geq s_k^T y_k \tau_k \rho_0 (\frac{6}{5}\max\{K, N_k, n_k\} - N_k) \\ &\geq s_k^T y_k \tau_k \rho_0 \frac{K}{5}. \end{aligned}$$

Then using the above inequality, we can transform (56) into

$$\begin{aligned} \|d_{k+1}\| &\leq \frac{5}{\rho_0 K s_k^T y_k \tau_k}(|q_1|\|g_{k+1}\| + |q_2|\|s_k\| + |q_3|\|y_k^*\|) \\ &\leq \frac{5}{\rho_0 K}\|g_{k+1}\|\left[\|g_{k+1}\|^2 + 4\frac{\|g_{k+1}\|^2\|y_k\|^2\|y_k^*\|^2}{s_k^T y_k \tau_k} + 2\frac{\|g_{k+1}\|^2\|s_k\|\|y_k\|}{s_k^T y_k}\right. \\ &\quad +4\frac{\|g_{k+1}\|\|s_k\|\|y_k\|\|y_k^*\||\omega_k|}{s_k^T y_k \tau_k} + 2\frac{\|g_{k+1}\|\|y_k^*\||\omega_k|}{\tau_k} + \frac{\rho_{k+1}\|s_k\|^2}{s_k^T y_k} \\ &\quad \left. + \frac{\|s_k\|^2\omega_k^2}{s_k^T y_k \tau_k} + 2\frac{\rho_{k+1}\|s_k\|\|y_k\|\|y_k^*\|^2}{s_k^T y_k \tau_k} + \frac{\rho_{k+1}\|y_k^*\|^2}{\tau_k}\right] \\ &\leq \frac{5\|g_{k+1}\|}{\rho_0}\frac{\|g_{k+1}\|^2}{K}\left(\frac{(2+\xi_2)^2}{\xi_1^2} + 6\frac{(2+\xi_2)}{\xi_1} + 4\frac{M_0}{\xi_1} + 2\sqrt{\frac{\xi_2}{\xi_1}} + 5\right) \\ &\leq \frac{\|g_{k+1}\|}{\rho_0 K_1}\left(\frac{25\xi_1^2 + 5\xi_2^2 + 30\xi_1\xi_2 + 10\xi_1\sqrt{\xi_1\xi_2} + (60 + 20M_0)\xi_1 + 20\xi_2 + 20}{\xi_1^2}\right). \end{aligned}$$

Since (26) implies that $K_1 \geq \xi_1$, we finally obtain an upper bound of d_{k+1}, that is

$$\|d_{k+1}\| \leq \left(\frac{25\xi_1^2 + 5\xi_2^2 + 30\xi_1\xi_2 + 10\xi_1\sqrt{\xi_1\xi_2} + (60 + 20M_0)\xi_1 + 20\xi_2 + 20}{\rho_0 \xi_1^3} \right) \|g_{k+1}\|.$$

For convenience, we define

$$L_2 \triangleq \left(\frac{25\xi_1^2 + 5\xi_2^2 + 30\xi_1\xi_2 + 10\xi_1\sqrt{\xi_1\xi_2} + (60 + 20M_0)\xi_1 + 20\xi_2 + 20}{\rho_0 \xi_1^3} \right),$$

namely, $\|d_{k+1}\| \leq L_2\|g_{k+1}\|$.

Case V (quadratic): if d_{k+1} is formed by (8) and (41), it is similar to Lemma 4 in Li, Liu and Liu [42]. According to their proof, we can deduce

$$\|d_{k+1}\| \leq \frac{20}{\vartheta_1}\|g_{k+1}\|.$$

If d_{k+1} is calculated by (12) and (40), it follows from [13] that

$$\|d_{k+1}\| \leq \frac{750}{\vartheta_1}\|g_{k+1}\|.$$

Since $\vartheta_1 > 0$, we finally have

$$\|d_{k+1}\| \leq \frac{750}{\vartheta_1}\|g_{k+1}\|.$$

According to the above analysis, the proof of Lemma 2 is completed by setting

$$\varsigma_2 = \max\left\{ \left(\frac{10\xi_2 + 5\xi_1 + 5m}{\bar{\rho}_0 \xi_1^2} \right), 1, 1 + \frac{L}{\vartheta_1}, L_2, \frac{750}{\vartheta_1} \right\}.$$

□

4.2. Convergence Analysis

In this subsection, we will give the global convergence of the presented algorithm for general functions.

Lemma 3. *Suppose α_k satisfies the line search conditions (5) and (6), and f satisfies Assumption 2, then*

$$\alpha_k \geq \frac{(1-\sigma)|g_k^T d_k|}{L\|d_k\|^2}. \tag{57}$$

Proof. According to the line search condition (5), we can easily obtain

$$(\sigma - 1)g_k^T d_k \leq (g_{k+1}^T - g_k)^T d_k = y_k^T d_k \leq \|y_k\|\|d_k\| \leq L\|s_k\|\|d_k\| = \alpha_k L\|d_k\|^2,$$

because $\sigma - 1 < 0$ and $g_k^T d_k < 0$, it is obvious that

$$\alpha_k \geq \frac{(1-\sigma)|g_k^T d_k|}{L\|d_k\|^2}.$$

□

Theorem 1. *Suppose the objective function f satisfies Assumptions 1 and 2, and sequence $\{x_k\}$ is generated by WYL_TSCO, then we have*

$$\lim_{k\to\infty} \|g_k\| = 0. \tag{58}$$

Proof. According to (57), (49) and (55), it follows

$$\delta\alpha_k g_k^T d_k \le -\frac{(1-\sigma)\delta}{L}\frac{(g_k^T d_k)^2}{\|d_k\|^2} \le -\frac{(1-\sigma)\delta\varsigma_1^2}{L\varsigma_2}\|g_k\|^2 = -H\|g_k\|^2,$$

where $H \triangleq \frac{(1-\sigma)\delta\varsigma_1^2}{L\varsigma_2}$, then combining (6), we have

$$f_{k+1} \le C_k + \delta\alpha_k g_k^T d_k \le C_k - H\|g_k\|^2. \tag{59}$$

With the reasonable assumptions that $Q_0 = 1$, $Q_1 = 2$, $Q_2 = 3$, $Q_3 = 4$, $Q_4 = 5$ and $\eta_0 = \eta_1 = \eta_2 = \eta_3 = \eta_4 = 1$ according to the update formulas (7) and (48), we can acquire that

$$Q_{k+1} = 1 + \sum_{j=0}^{k}\prod_{i=0}^{j}\eta_{k-i}, \quad k = 0, 1, 2...$$

Combining (48), we can derive the general formula of Q_{k+1},

$$Q_{k+1} = \begin{cases} 1 + \eta^{k/n} + n\sum_{i=1}^{k/n}\eta^i, & \text{mod } (k,n) = 0, \\ 1 + \text{mod}(k,n) + \eta^{\lfloor k/n \rfloor} + n\sum_{i=1}^{\lfloor k/n \rfloor}\eta^i, & \text{mod } (k,n) \neq 0, \end{cases}$$

where $\lfloor\cdot\rfloor$ denotes the floor function. Thus, it is easy to obtain an upper bound of Q_{k+1},

$$\begin{aligned} Q_{k+1} &\le 1 + \text{mod}(k,n) + \eta^{\lfloor k/n \rfloor + 1} + n\sum_{i=1}^{\lfloor k/n \rfloor + 1}\eta^i \\ &\le 1 + n + (n+1)\sum_{i=1}^{\lfloor k/n \rfloor + 1}\eta^i \\ &\le 1 + n + (n+1)\sum_{i=1}^{k+1}\eta^i \\ &\le (1+n)\sum_{i=0}^{k+1}\eta^i \\ &\le \frac{1+n}{1-\eta} \\ &= W, \end{aligned} \tag{60}$$

where $W = \frac{1+n}{1-\eta}$.

When $k \ge 5$, combining (7), (59) and (60), we obtain

$$C_{k+1} = C_k + \frac{f_{k+1} - C_k}{Q_{k+1}} \le C_k - \frac{H}{W}\|g_k\|^2,$$

which means

$$\frac{H}{W}\|g_k\|^2 \le C_k - C_{k+1}. \tag{61}$$

When $k < 5$, (6) and (46) imply that

$$C_{k+1} \le f_{k+1} + 0.9(C_k - f_{k+1}) = C_k + 0.1(f_{k+1} - C_k) < C_k.$$

Hence C_k is monotonically decreasing. According to Lemma 1.1 in [15], we have $f_{k+1} \le C_{k+1}$ for each $k \ge 5$, and (46) indicates that $f_{k+1} \le C_{k+1}$ for each $k < 5$. Thus, with the assumption that f is bounded below, C_k is certainly bounded from below.

Summing up the above analysis and (61), we can finally obtain that

$$\sum_{k=0}^{\infty} \frac{H}{W} \|g_k\|^2 < \infty,$$

which implies

$$\lim_{k\to\infty} \|g_k\|^2 = 0.$$

The proof is completed. □

5. Numerical Results

In this section, the results of the numerical experiments are shown below. The unconstrained test functions were taken from [62] with the given initial points. To prove the efficiency of the proposed WYL_TSCO algorithm, we compare its numerical performance with SMCG_Conic and TSCG. The performance profile proposed by Dolan and Moré [63] is used to evaluate the performance of these methods. The dimension of the test functions is 10,000. All the programs were written in C code.

SMCG_Conic is also a subspace minimization conjugate gradient algorithm based on the conic model, and is a pioneering one that combines the subspace technique with a conic model to seek the search direction. The numerical experiments in [52] show that the performance of SMCG_Conic is very efficient. Since the biggest difference between WYL_TSCO and SMCG_Conic is the dimension of the used subspace, the comparison between WYL_TSCO and SMCG_Conic can not only reflect the high efficiency of our algorithm, but also reveal the influence on the numerical result generated by the change of dimension for the adopted subspace in the subspace minimization conjugate gradient algorithm.

As for TSCG, it is an efficient subspace minimization conjugate gradient method, and it successfully applies the idea of subspace method and WYL method which are employed in our algorithm as well. However, TSCG only uses the quadratic model to approximate the objective function. Thus, the comparison between the numerical performance of WYL_TSCO and TSCG can reveal the influence of the approximation model in the subspace method.

For the initial stepsize of the first iteration, we adopt the adaptive strategy used in [42]. The other parameters of WYL_TSCO are selected as follows.

$\epsilon = 10^{-6},\ \epsilon_1 = 10^{-3},\ \epsilon_2 = 10^{-4},\ \delta = 0.001,\ \sigma = 0.9999,\ \lambda_{min} = 10^{-30},\ \lambda_{max} = 10^{30},$
$\xi_1 = 0.15 \times 10^2,\ \xi_2 = 8.5 \times 10^4,\ \xi_3 = 4 \times 10^8,\ \xi_4 = 6.5 \times 10^7,\ \xi_5 = 0.1,\ \rho_0 = 0.3,$
$\bar{\rho}_0 = 0.9,\ \vartheta_1 = 10^{-7},\ \vartheta_2 = 10^4,\ \vartheta_3 = 10^{-2},\ \vartheta_4 = 10^{-4},\ \lambda_1 = 10^{-7},\ \lambda_2 = 0.05.$

SMCG_Conic and TSCG use the original parameters in their paper respectively. In addition to stopping when the stopping criterion $\|g_k\|_\infty \leq \varepsilon$ holds, the algorithm also stops when the number of iterations exceeds 200,000.

For the 80 test problems from [62], WYL_TSCO successfully solves 100% of them, while TSCG and SMCG_Conic both successfully solve 76 problems. We compare the numerical results of problems that are solved successfully by all the three algorithms, the number of which are 72. The numerical results of the three algorithms are presented in the following figures.

Figure 1 depicts the performance based on the number of iterations for the three methods. It shows that the performance of WYL_TSCO is similar to that of TSCG, but is a little inferior to SMCG_Conic's only when $\tau < 1.31$. However, WYL_TSCO outperforms TSCG and SMCG_Conic when $\tau > 1.31$.

In Figure 2, we observe that WYL_TSCO and TSCG outperforms SMCG_Conic on the number of function evaluations, and TSCG is slightly ahead of WYL_TSCG only in the case of $\tau < 1.5$.

Similar to Figure 1, Figure 3 illustrates that WYL_TSCO is at a little disadvantage compared with SMCG_Conic only when $\tau < 1.318$, but is competitive with TSCG.

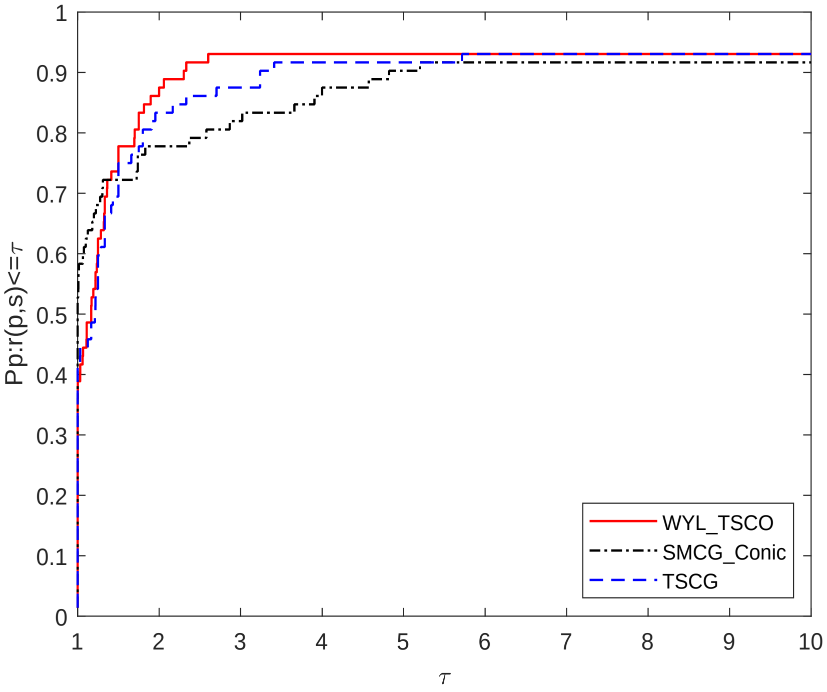

Figure 1. Performance profile based on the number of iterations.

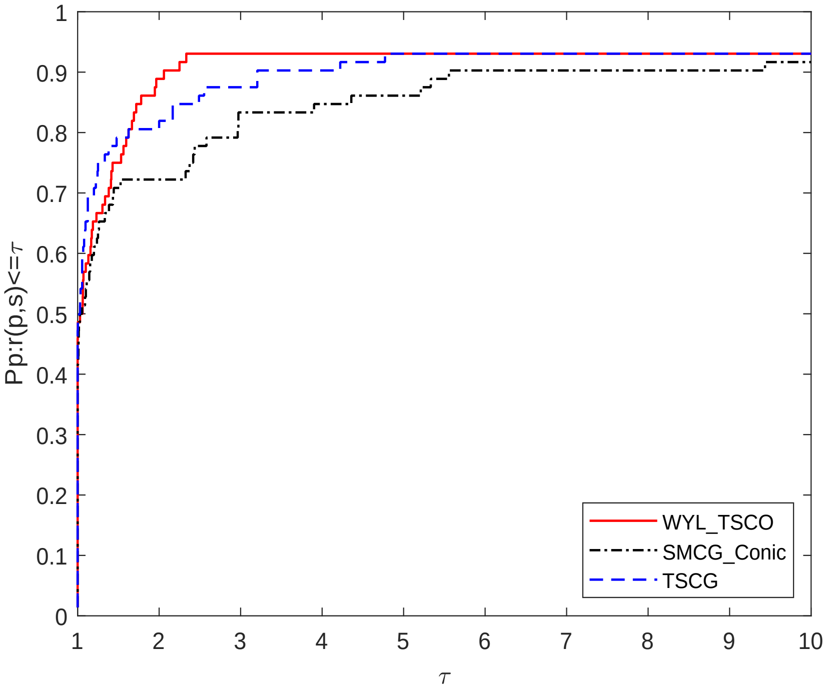

Figure 2. Performance profile based on the number of function evaluations.

For the CPU time, Figure 4 shows that WYL_TSCO has an improvement on TSCG and SMCG_Conic, which shows the high efficiency of WYL_TSCO.

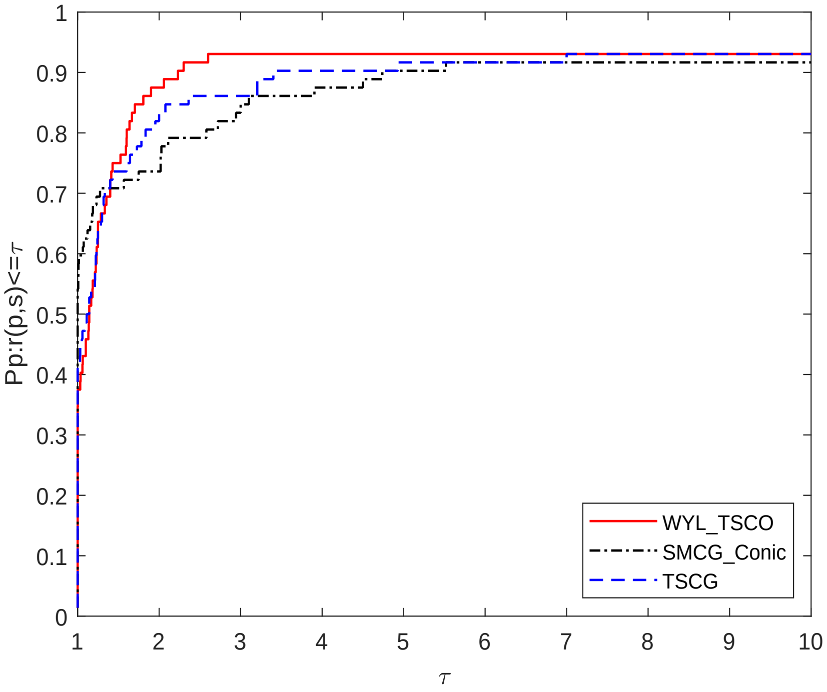

Figure 3. Performance profile based on the number of gradient evaluations.

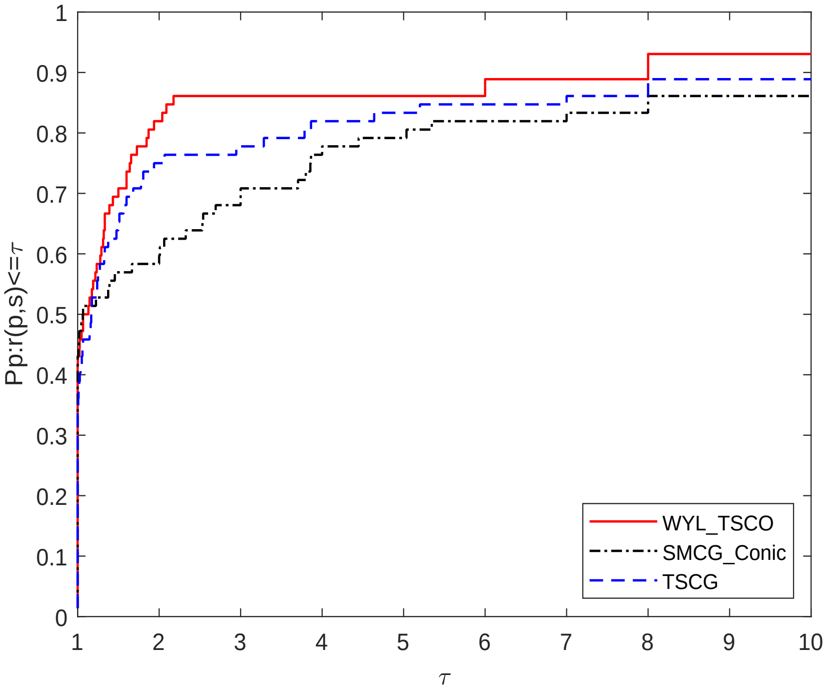

Figure 4. Performance profile based on the CPU time.

For the 80 test problems, the numerical results show that WYL_TSCO and TSCG may perform a little worse than SMCG_Conic when τ is very small, but they overall have significant improvements over SMCG_Conic, and WYL_TSCO also is superior to TSCG for almost all τ. Moreover, WYL_TSCO can solve more problems than TSCG and SMCG_Conic, which might show the advantage of the three-dimensional subspace method and conic model. To sum up, the WYL_TSCO is an efficient algorithm for solving the unconstrained optimization problem.

6. Conclusions

Combining the idea of WYL method, this paper proposes a three-dimensional subspace method based on conic models for unconstrained optimization (WYL_TSCO). The sufficient descent property of the search direction and the global convergence of this method are obtained under some suitable assumptions. The selection of an approximation model depends on whether certain criterions are satisfied. Furthermore, the strategies of initial stepsize and nonmonotone line search are exploited which are beneficial to the convergence and efficiency. According to the numerical results and theoretical analysis, WYL_TSCO is very competitive and promising.

Author Contributions: Conceptualization, S.Y.and J.X.; methodology, G.W.; software and validation, G.W. and M.P.; visualization and formal analysis, G.W.; writing—original draft preparation, G.W.; supervision, S.Y.; funding acquisition, S.Y. and J.X. All authors have read and agreed to the published version of the manuscript.

Funding: Chinese National Natural Science Foundation no. 12262002, Natural Science Foundation of Guangxi Province (CN) no. 2020GXNSFAA159014 and 2021GXNSFAA196076, the Program for the Innovative Team of Guangxi University of Finance and Economics, and Guangxi First-class Discipline statistics Construction Project.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hestenes, M.R.; Stiefel, E.L. Methods of conjugate gradients for solving linear systems. *J. Res. Natl. Bur. Stand.* **1952**, *49*, 409–436. [CrossRef]
2. Fletcher, R.; Reeves, C.M. Function minimization by conjugate gradients. *Comput. J.* **1964**, *7*, 149–154. [CrossRef]
3. Polyak, B.T. The conjugate gradient method in extremal problems. *Jussr Comput. Math. Math. Phys.* **1969**, *9*, 94–112. [CrossRef]
4. Fletcher, R. Volume 1 Unconstrained Optimization. In *Practical Methods of Optimization*; Wiley: New York, NY, USA, 1987.
5. Liu, Y.L.; Storey, C.S. Efficient generalized conjugate gradient algorithms, part 1: Theory. *J. Optim. Theory Appl.* **1991**, *69*, 129–137. [CrossRef]
6. Dai, Y.H.; Yuan, Y.X. A nonlinear conjugate gradient method with a strong global convergence property. *Siam J. Optim.* **1999**, *10*, 177–182. [CrossRef]
7. Wei, Z.; Yao, S.; Liu, L. The convergence properties of some new conjugate gradient methods. *Appl. Math. Comput.* **2006**, *183*, 1341–1350. [CrossRef]
8. Huang, H.; Wei, Z.; Yao, S. The proof of the sufficient descent condition of the Wei-Yao-Liu conjugate gradient method under the strong Wolfe-Powell line search. *Appl. Math. Comput.* **2007**, *189*, 1241–1245. [CrossRef]
9. Zhang, L. An improved Wei-Yao-Liu nonlinear conjugate gradient method for optimization computation. *Appl. Math. Comput.* **2009**, *215*, 2269–2274. [CrossRef]
10. Yao, S.; Qin, B. A hybrid of DL and WYL nonlinear conjugate gradient methods. *Abstr. Appl. Anal.* **2014**, *2014*, 279891. [CrossRef]
11. Huang, H.; Lin, S. A modified Wei-Yao-Liu conjugate gradient method for unconstrained optimization. *Appl. Math. Comput.* **2014**, *231*, 179–186. [CrossRef]
12. Hu, Y.; Wei, Z. Wei-Yao-Liu conjugate gradient projection algorithm for nonlinear monotone equations with convex constraints. *Int. J. Comput. Math.* **2015**, *92*, 2261–2272. [CrossRef]
13. Huo, J.; Yang, J.; Wang, G.; Yao, S. A class of three-dimensional subspace conjugate gradient algorithms for unconstrained optimization. *Symmetry.* **2022**, *14*, 80. [CrossRef]
14. Wolfe, P. Convergence conditions for ascent methods. *SIAM Rev. Soc. Ind. Appl. Math.* **1969**, *11*, 226–235. Erratum in *SIAM Rev. Soc. Ind. Appl. Math.* **1971**, *13*, 185–188. [CrossRef]
15. Zhang, H.; Hager, W.W. A nonmonotone line search technique and its application to unconstrained optimization. *SIAM J. Optim.* **2004**, *14*, 1043–1056. [CrossRef]
16. Grippo, L.; Lampariello, F.; Lucidi S. A nonmonotone line search technique for Newton's method. *SIAM J. Numer. Anal.* **1986**, *23*, 707–716. [CrossRef]

17. Hager, W.W.; Zhang, H. A new conjugate gradient method with guaranteed descent and an efficient line search. *SIAM J. Optim.* **2005**, *16*, 170–192. [CrossRef]
18. Gu, N.Z.; Mo, J.T. Incorporating nonmonotone strategies into the trust region method for unconstrained optimization. *Comput. Math. Appl.* **2008**, *55*, 2158–2172. [CrossRef]
19. Dai, Y.H.; Kou, C.X. A nonlinear conjugate gradient algorithm with an optimal property and an improved Wolfe line search. *SIAM J. Optim.* **2013**, *23*, 296–320. [CrossRef]
20. Huang, S.; Wan, Z.; Chen, X. A new nonmonotone line search technique for unconstrained optimization. *Numer. Algorithms.* **2015**, *68*, 671–689. [CrossRef]
21. Ou, Y.; Liu, Y. A memory gradient method based on the nonmonotone technique. *J. Ind. Manag. Optim.* **2017**, *13*, 857–872. [CrossRef]
22. Ahookhosh, M.; Amini, K.; Peyghami, M.R. A nonmonotone trust-region line search method for large-scale unconstrained optimization. *Appl. Math. Model.* **2012**, *36*, 478–487. [CrossRef]
23. Ahookhosh, M.; Amini, K. An efficient nonmonotone trust-region method for unconstrained optimization. *Numer. Algorithms.* **2012**, *59*, 523–540. [CrossRef]
24. Kimiaei, M.; Rahpeymaii, F. A new nonmonotone line-search trust-region approach for nonlinear systems. *TOP* **2019**, *27*, 199–232. [CrossRef]
25. Kimiaei, M. A new class of nonmonotone adaptive trust-region methods for nonlinear equations with box constraints. *Calcolo.* **2017**, *54*, 769–812. [CrossRef]
26. Yuan, Y.X. Subspace techniques for nonlinear optimization. In *Some Topics in Industrial and Applied Mathematics*; Jeltsch, R., Li, D.Q., Sloan, I.H., Eds.; Higher Education Press: Beijing, China, 2007; pp. 206–218.
27. Yuan, Y.X. A review on subspace methods for nonlinear optimization. In Proceedings of the International Congress of Mathematics, Seoul, Republic of Korea, 13–21 August, 2014; pp. 807–827.
28. Liu, D.C.; Nocedal, J. On the limited memory BFGS method for large scale optimization. *Math. Program.* **1989**, *45*, 503–528. [CrossRef]
29. Wang, Z.H.; Wen, Z.W.; Yuan, Y. A subspace trust region method for large scale unconstrained optimization. In *Numerical Linear Algebra and Optimization*; Yuan, Y., Ed.; Science Press: Beijing, China, 2004; pp. 265–274.
30. Wang, Z.H.; Yuan, Y. A subspace implementation of quasi-Newton trust region methods for unconstrained optimization. *Numer. Math.* **2006**, *104*, 241–269. [CrossRef]
31. Lee, J.H.; Jung, Y.M.; Yuan, Y.X.; Yun, S. A subspace SQP method for equality constrained optimization. *Comput. Optim. Appl.* **2019**, *74*, 177–194. [CrossRef]
32. Gill, P.E.; Leonard, M.W. Reduced-hessian quasi-newton methods for unconstrained optimization. *SIAM J. Optim.* **2001**, *12*, 209–237. [CrossRef]
33. Liu, X.; Wen, Z.; Zhang, Y. Limited memory block Krylov subspace optimization for computing dominant singular value decompositions. *SIAM J. Sci. Comput.* **2013**, *35*, A1641–A1668. [CrossRef]
34. Yuan, Y.X. Subspace methods for large scale nonlinear equations and nonlinear least squares. *Optim. Eng.* **2009**, *10*, 207–218. [CrossRef]
35. Dong, Q.; Liu, X.; Wen, Z.W.; Yuan, Y.X. A parallel line search subspace correction method for composite convex optimization. *J. Oper. Res. Soc. China.* **2015**, *3*, 163–187. [CrossRef]
36. Li, Y.; Liu, H.; Wen, Z.; Yuan, Y.X. Low-rank matrix iteration using polynomial-filtered subspace extraction. *SIAM J. Sci. Comput.* **2020**, *42*, A1686-A1713. [CrossRef]
37. Kimiaei, M.; Neumaier, A.; Azmi, B. LMBOPT: A limited memory method for bound-constrained optimization. *Math. Prog. Comp.* **2022**, *14*, 271–318. [CrossRef]
38. Cartis, C.; Roberts, L. Scalable subspace methods for derivative-free nonlinear least-squares optimization. *Math. Program.* **2023**, *199*, 461–524. [CrossRef]
39. Yuan, Y.X.; Stoer, J. A subspace study on conjugate gradient algorithms. *Z. Angew. Math. Mech.* **1995**, *75*, 69–77. [CrossRef]
40. Andrei, N. An accelerated subspace minimization three-term conjugate gradient algorithm for unconstrained optimization. *Numer. Algorithms.* **2014**, *65*, 859–874. [CrossRef]
41. Yang, Y.; Chen, Y.; Lu, Y. A subspace conjugate gradient algorithm for large-scale unconstrained optimization. *Numer. Algorithms.* **2017**, *76*, 813–828. [CrossRef]
42. Li, M.; Liu, H.; Liu, Z. A new subspace minimization conjugate gradient method with nonmonotone line search for unconstrained optimization. *Numer. Algorithms.* **2018**, *79*, 195–219. [CrossRef]
43. Dai, Y.H.; Kou, C.X. A Barzilai-Borwein conjugate gradient method. *Sci. China Math.* **2016**, *59*, 1511–1524. [CrossRef]
44. Barzilai, J.; Borwein, J.M. Two-point step size gradient methods. *IMA J. Numer. Anal.* **1988**, *8*, 141–148. [CrossRef]
45. Sun W. On nonquadratic model optimization methods. *Asia. Pac. J. Oper. Res.* **1996**, *13*, 43–63.
46. Sun W.; Yuan Y. *Optimization Theory and Methods: Nonlinear Programming*; Springer: New York, NY, USA, 2006.
47. Davidon, W.C. Conic Approximations and Collinear Scalings for Optimizers. *SIAM J. Numer. Anal.* **1980**, *17*, 268–281. [CrossRef]
48. Sorensen, D.C. The Q-superlinear convergence of a collinear scaling algorithm for unconstrained optimization. *SIAM J. Numer. Anal.* **1980**, *17*, 84–114. [CrossRef]
49. Ariyawansa, K.A. Deriving collinear scaling algorithms as extensions of quasi-Newton methods and the local convergence of DFP- and BFGS-related collinear scaling algorithms. *Math. Program.* **1990**, *49*, 23–48. [CrossRef]
50. Sheng, S. Interpolation by conic model for unconstrained optimization. *Computing.* **1995**, *54*, 83–98. [CrossRef]

51. Di, S.; Sun, W. A trust region method for conic model to solve unconstraind optimizaions. *Optim. Methods Softw.* **1996**, *6*, 237–263. [CrossRef]
52. Li, Y.; Liu, Z.; Liu, H. A subspace minimization conjugate gradient method based on conic model for unconstrained optimization. *Comput. Appl. Math.* **2019**, *38*, 16. [CrossRef]
53. Sun, W.; Li, Y.; Wang, T.; Liu, H. A new subspace minimization conjugate gradient method based on conic model for large-scale unconstrained optimization. *Comput. Appl. Math.* **2022**, *41*, 178. [CrossRef]
54. Yuan, Y.X. A modified BFGS algorithm for unconstrained optimization. *IMA J. Numer. Anal.* **1991**, *11*, 325–332. [CrossRef]
55. Dai, Y.H.; Yuan, J.Y.; Yuan, Y.X. Modified two-point stepsize gradient methods for unconstrained optimization problems. *Comput. Optim. Appl.* **2002**, *22*, 103–109. [CrossRef]
56. Liu, Z.; Liu, H. An efficient gradient method with approximate optimal stepsize for large-scale unconstrained optimization. *Numer. Algorithms.* **2018**, *78*, 21–39. [CrossRef]
57. Liu, H.; Liu, Z. An efficient Barzilai-Borwein conjugate gradient method for unconstrained optimization. *J. Optim. Theory Appl.* **2019**, *180*, 879–906. [CrossRef]
58. Nocedal, J.; Wright, S.J. *Numerical Optimization*; Springer: New York, NY, USA, 2006.
59. Andrei, N. *Nonlinear Conjugate Gradient Methods for Unconstrained Optimization*; Springer: Cham, Switzerland, 2020.
60. Hager, W.W.; Zhang, H. Algorithm 851: CG_DESCENT, a conjugate gradient method with guaranteed descent. *ACM Trans. Math. Softw.* **2006**, *32*, 113–137. [CrossRef]
61. Andrei, N. Open problems in nonlinear conjugate gradient algorithms for unconstrained optimization. *Bull. Malays. Math. Sci. Soc. Second Series.* **2011**, *34*, 319–330.
62. Andrei, N. An unconstrained optimization test functions collection. *Adv. Model. Optim.* **2008**, *10*, 147–161.
63. Dolan, E.D.; Moré, J.J. Benchmarking optimization software with performance profiles. *Math. Program.* **2002**, *91*, 201–213. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.